PA 1193958

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

July 13, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/507,488***
FILING DATE: ***October 02, 2003***

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

L. Edelen

L. EDELEN
Certifying Officer

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

10/06/2003 AADOF01 00000107 501407 60507488
01 FC:2005 80.00 DA

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002 — 489-267/69033

Approved for use through 01/31/98. OMB 0651-0037
Patent and Trademark Office; US DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a **PROVISIONAL APPLICATION FOR PATENT** under 37 CFR 1.53 (b)(2).

Docket Number	26919	Type a plus sign (+) inside this box ->	+

INVENTOR(s) / APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
GAZIT	Ehud		Ramat Hasharon, Israel
CHERNY	Izhack		Petah-Tikva, Israel

TITLE OF THE INVENTION (280 characters max)

ANTI BACTERIAL TARGETS AND AGENTS AND METHODS OF IDENTIFYING SAME

CORRESPONDENCE ADDRESS

G. E. EHRLICH (1995) LTD.
c/o ANTHONY CASTORINA
2001 JEFFERSON DAVIS HIGHWAY
SUITE 207

STATE	VIRGINIA	ZIP CODE	22202	COUNTRY	USA

ENCLOSED APPLICATION PARTS *(check all that apply)*

☑ Specification	*Number of Pages*	29	☑ Applicant is entitled to Small Entity Status	
☑ Drawing(s)	*Number of Sheets*	7	☑ Other (specify) *ASSIGNMENT*	

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☐ A check or money order is enclosed to cover the filing fees		**FILING FEE AMOUNT ($)**	**$ 80.-**
☑ The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number:	**50-1407**		

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No

☐ Yes, the name of the US Government agency and the Government contract number are: _____

Respectfully submitted,

SIGNATURE _____ Date 30 September 2003

TYPED or PRINTED NAME SOL SHEINBEIN

REGISTRATION NO. (*if appropriate*) 25,457

☐ Additional inventors are being named on separately numbered sheets attached hereto

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

ANTI BACTERIAL TARGETS AND AGENTS AND METHODS OF IDENTIFYING SAME

Inventors: Ehud Gazit [and] Izhack Cherny

Abbreviations used: CD, circular dichroism; FTIR, Fourier Transform Infrared; GST, glutathione s-transferase; ORF, open reading frame; TA, Toxin-Antitoxin.

Abstract

While natively unfolded proteins are being increasingly observed, their physiological role is not well understood. Here, we demonstrate that the *Escherichia coli* YefM protein is a natively unfolded antitoxin, lacking secondary structure even at low temperature or in the presence of stabilizing agent. This conformation of the protein is suggested to have a key role in its physiological regulatory activity. Due to the unfolded state of the protein, a linear determinant rather than a conformational one is presumably being recognized by its toxin-partner, YoeB. A peptide array technology allowed the identification and validation of such a determinant. This recognition element may provide a novel antibacterial target. Indeed, a pair-constrained bioinformatics analysis facilitated the definite determination of novel YefM-YoeB toxin-antitoxin systems in a large number of bacteria including major pathogens such as *Staphylococcus aureus*, *Streptococcus pneumoniae*, and *Mycobacterium tuberculosis*. Taken together, the YefM protein defines a new family of natively unfolded proteins. The existent of a large and conserved group of proteins with a clear physiologically- relevant unfolded state serves as a paradigm to understand the structural basis of this state.

Introduction

The "thermodynamic hypothesis" of protein folding, as was introduced more than forty years ago, suggests that the folded state of a given protein represents a global minimum of free energy [1]. While this theory is widely valid, there is a considerable group of "natively unfolded" proteins (as were first denoted by Mandelkow and coauthors [2]) that rather favors the thermodynamically unfolded state [3-6]. For a recent review on natively unfolded proteins see Uversky, 2002 [5]. The unfolded state of this group of proteins does not signify a requirement for the activity of molecular chaperons to overcome a large energetic barrier to attain a global minimum energy, but a truly energetically favorable unfolded state. The natively unfolded state is also distinct from the misfolded state in which proteins self-assemble to form large supramolecular assemblies such as amyloid fibrils [7-9].

While the number of natively unfolded proteins identified is steady increasing [4,10], their physiological significance is poorly understood. One case in which a natively unfolded state of a protein appears to have physiological significance is that of the Phd protein of the phage P1 [11]. This protein is a part of a bimolecular complex that acts as the "plasmid addiction" module of the phage [12]. The addiction module mechanism assures an efficient inheritance of the extrachromosomal phage and is based on the differential physiological stability of its two components, the stable toxin Doc and the labile antitoxin Phd. Upon a loss of the phage in a postsegregational event, no *de novo* synthesis of either the toxin or antitoxin occurs. Due to the physiological instability of the antitoxin, only the toxic component of the module is ultimately retained within the cured cells, causing the death of cured cells. Consistent with the fact that Phd is recognized and degraded by the ClpXP "quality control"

machinery of infected cells [13], we suggested that its unfolded state is the key to its physiological instability, thus serving as a critical element in the function of the TA module. Many "damaged" or misfolded proteins are identified and eliminated by the ClpXP system. These unfolded target proteins may be recognized by ectopic exposure of hydrophobic amino acids, which are normally buried within the hydrophobic core of the protein. Therefore, we assumed that ClpXP recognizes the unfolded Phd protein based on its structural property, as it may appear as damaged protein.

TA systems were also identified on chromosomes in both bacteria and archea, but not in eurokaryotes [14-19]. These systems share the same paradigm of a stable toxin and an unstable antidote, organization as a polycistronic operon, and small size of the protein components (70-100 amino acids). Although TA systems are widely present, their physiological role is not fully understood. It is assumed that the systems play a significant role in survival under stringent conditions [14-19].

The absolute lack of TA systems in euokaryotes, as oppose to their ubiquitous presence in bacteria and archea, makes the systems a very attractive antibacterial target. Unlike conventional antibiotics, there is no need for the external introduction of toxic material that may affect the host as well. The blockage of the toxin-antitoxin physical interaction may result in the execution of the inherent toxic potential of the toxin.

In this work, we clearly demonstrate that the *E. coli* YefM antitoxin protein, although showing very low homology to the Phd protein, is also natively unfolded. Pair-constrained bioinformatics analysis allowed the identification of a large family of

natively unfolded host proteins that are based on the Phd-YefM structural framework. The chromosomal organization of the proteins implies that they are a part of functional TA systems in a related group of bacteria, including some major pathogens. The unfolded YefM-like proteins are attractive target for the development of antibacterial agents, due to the fact that the toxin partner of the TA module recognizes a linear determinant with the antitoxin, which could be mimicked by a therapeutic agent.

Experimental Procedures

Genes Sequence identification and alignments

Sequences related to the *yefM* and *yoeB* genes of *E. coli* were identified by a pair-constrained bioinformatical analysis. Sequences were identified using TBLASTN and PSI-BLAST searches [20] of nonredundant microbial genomes database at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). Putative *yefM* and *yoeB* homologue sequences were obtained and examined for constituting a toxin-antitoxin gene-pair module in the chromosome. Low homology unpaired sequences were discarded. Alignments were produced by CLASTAL W [21] with default settings and edited using JALVIEW editor.

Cloning of the system genes into pBAD-TOPO expression vector

DNA fragments containing the coding sequence of *yefM*, *yoeB* and both *yefM-yoeB*, were produced by PCR using the chromosomal DNA of *E. coli* K-12 MC1061, and the primers ATGYEFM (5'-ATGAACTGTACAAAAGAGG-3') and YEFMEND (5'-GACAAGCTTAGTTTCACTCAATG-3') to amplify *yefM* gene; GTGYOEB (5'-GTGAAACTAATCTGGTCTG-3') and YOEBEND1 (5'-

TGAAGCTTTTCAATAATGATAACGAC-3') to amplify *yoeB* gene; and ATGYEFM and YOEBEND1 to amplify *yefM-yoeB* genes together. The PCR fragments, using the pBAD-TOPO TA cloning kit (invitrogen), were cloned into the pBAD-TOPO vector to generate pBAD-yefM, pBAD-yoeB, and pBAD-yefMyoeB. The plasmids were transformed into an *E. coli* TOP10 strain (Invitrogen).

Growth rate analysis

E. coli TOP10 bacteria transformed with pBAD-yefM, pBAD-yoeB, and pBAD-yefMyoeB were cultured overnight in LB broth supplemented with 100 μg/ml ampicillin at 37 °C. On the next day, the three cultures were diluted and adjusted to optical density of approximately 0.01 (A_{600}) in LB-Amp. Next, each culture was divided into two equal volumes, whereas, at time zero, the first half was added with 0.2% L-arabinose to induce expression of the target gene and the second half with 0.2% D-glucose to suppress low transcription from the pBAD promoter. All cultures were grown at 37 °C/200 rpm, and samples were sequentially taken approximately every 40-60 minutes for 9 hours. Cells density was measured by its optical absorbance at 600 nm. To inspect growth rate for gene induction during logarithmic growth phase, the same analysis assay as above was conducted, with the exception of the time of induction. Cultures were divided and expression was induced (or suppressed) at the time they had reached optical density of approximately 0.45 (A_{600}).

Colony formation analysis

E. coli TOP10 bacteria transformed with pBAD-yefM, pBAD-yoeB, and pBAD-yefMyoeB, were grown in LB broth at 37 °C containing ampicillin as indicated. After overnight growth, cultures were diluted to an A_{600} of 0.01 in LB-Amp medium. The cultures were then grown at 37 °C until an A_{600} value of 0.5 was reached. At that point, cells were diluted 10^4 to 10^7 times in ten-fold dilutions steps, and applied as 5 µl dropouts on LB-amp-agar plates containing arabinose in the following decreasing arabinose dilutions: 0.2%, 0.1%, 0.05%, 0.02%, 0.005% and 0.0005%. In addition, a negative control plate without arabinose and supplemented with 0.2% glucose was plated. All plates were incubated at 37 °C for at least 20 hours.

Cloning, expression and purification of YefM from E. coli

The DNA fragment containing the coding sequence of *yefM,* flanked by primer-encoded *BsrGI* and *HindIII* sites, was produced by a polymerase chain reaction using *E. coli* K-12 MC1061 strain chromosome as template and oligonucleotide primers YEFMSTART (5'-GTACAATGAACTGTACAAAAGAAG-3') and YEFMEND (5'-GACAAGCTTAGTTTCACTCAATG-3'). The product was digested with *BsrGI* and *HindIII* enzymes (New England Biolabs), cloned into the *BsrGI* and *HindIII* restriction sites of a pET42a expression vector (Novagen) in fusion to glutathione s-transferase (GST) and transformed into *E. coli* BL21(DE3) pLysS (Novagen). Transformed bacteria were grown in 2YT broth at 37 °C/200 rpm to an optical density (A_{600}) of approximately 0.4. Protein expression was induced by the addition of IPTG (2 mM). After 1 hour, cells were harvested and resuspended in phosphate buffer saline pH 7.3 (PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4), protease inhibitor cocktail as recommended (Sigma), and 0.5 mM PMSF,

and lysed by three passages through a French-press cell (1400 psi). The insoluble material was removed by centrifugation for 20 min at 20,000 x *g* at 4 °C, followed by a 0.45-μm filtration. The supernatant was applied onto a 1 ml glutathione sepharose column (Amersham Pharmacia Biotech) pre-equilibrated with PBS pH 7.3. The protein was eluted using 10 ml of 50 mM Tris-HCl (pH 8.0), 10 mM glutathione. YefM proteins were separated from the GST using 16 units of factor Xa protease (Novagen) per 1 mg YefM fusion. After 14 hours incubation at 37 °C, reaction was terminated by the addition of 1mM PMSF. Two different methods were applied for YefM purification. In the first method, gel filtration was conducted in order to remove the GST and linker protein (~40 kDa) from YefM (~11 kDa) using a Sepharose HR 10/30 (FPLC) gel filtration column (Amersham Pharmacia Biotech) and a FPLC instrument (Pharmacia LBK). Proteins were eluted with PBS pH 7.3, 0.8 ml/min, and a peak that included the ~11 kDa YefM proteins was collected after 13 min. Fractions containing the YefM protein were completely purified using 1 µmol of immobilized glutathione agarose (Sigma) agitated for 16 hours at room temperature. At this point, YefM was greater than 95% pure as estimated by Coomassie staining of SDS-PAGE. In the second purification method, the YefM and GST proteins mixture was divided into 0.5 ml fractions, boiled for 10 minutes and then centrifuged at 14,000 rpm for 10 minutes. The supernatants, containing the purified YefM, were collected and united. In order to determine YefM concentration, tyrosine absorbance measurement in 0.1M KOH was used. Protein concentrations were calculated using the extinction coefficients of 2391 M^{-1} cm^{-1} (293.2 nm in 0.1 M KOH) for single tyrosine.

The molecular mass of YefM was verified by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry using a voyager-DE STR

Biospectrometry workstation (Applied Biosystems). α-Cyano-4-hydroxycinnamic acid was used as the matrix.

Cloning, expression, and purification of GST-YoeB from E. coli

The DNA fragment containing the coding sequence of *yoeB,* flanked by primer-encoded *EcoRI* and *HindIII* sites, was produced by a polymerase chain reaction using *E. coli* K-12 MC1061 strain chromosome as template and oligonucleotide primers YOEBSTART (5'-AAAGGACATGAATTCGTGAAACTAATC-3') and YOEBEND2 (5'- CCTTTGAAGCTTTTCAATAATGATAA-3'). The product was digested with *EcoRI* and *HindIII* enzymes (New England Biolabs), cloned into the *EcoRI* and *HindIII* restriction sites of the pET42a expression vector in fusion to GST, and transformed into *E. coli* BL21(DE3) pLysS. Bacteria were grown, expressed and lysed in the same manner described above for GST-YefM fusion. The supernatant was applied onto a 1 ml glutathione sepharose column (Amersham Pharmacia Biotech) pre-equilibrated with PBS pH 7.3. The bound protein was eluted using 10 ml of 50 mM Tris-HCl (pH 8.0), 10 mM glutathione. Eluted fractions containing the GST-YoeB protein were collected and quantitatively assessed by Coomassie staining of SDS-PAGE.

Circular Dichroism (CD)

CD spectra were obtained by using an AVIV 202 spectrapolarimeter equipped with temperature-controlled sample holder and a 5 mm path length cuvette. Mean residual ellipticity, $[\theta]$, was calculated as,

$$[\theta] = [100 \times \theta \times m] / [c \times L]$$

where θ is the observed ellipticity, m is the mean residual weight, c is the concentration in mg/ml, and L is the path length in centimeters. All experiments were performed in PBS pH 7.3, at protein concentration of 10 μM. For thermal denaturation experiments, samples were equilibrated at each temperature for 0.5 min, and CD ellipticity at 222 nm and 217 nm was averaged for 1 min.

Fourier Transform Infrared Spectroscopy

Infrared spectra were recorded using a Nicolet Nexus 470 FT-IR spectrometer with a DTGS detector. The sample, 1μg of lyophilized YefM suspended in 30μl PBS in D_2O pD 7.3, was suspended on a CaF_2 plate. The measurements were taken using a 4 cm^{-1} resolution and 2,000 scans averaging. The transmittance minima values were determined by the OMNIC analysis program (Nicolet).

Amino acid composition and charge-hydrophobicity values analysis

The rate of occurrence of each amino acid in the YefM family proteins (P_{Mi}) was determined by averaging its 30 frequencies in each of the 30 YefM homologue sequences. The general amino acid occurrence statistics (P_{Gi}) were compiled by the Rockerfeller authors using the NCBI database [22]. The comparison ordinates between the amino acids occurrences are given by their fractional difference: $(P_{Mi} - P_{Gi}) / P_{Gi}$. The variances of these ratios were calculated as: $\mathrm{Var}(P_{Mi}) / (P_{Gi})^2$.

The mean hydrophobicity and the mean net charge of the YefM and the YefM homologues proteins were calculated as described by Uversky and coauthors [3].

Peptide array analysis

Tridecamer peptides corresponding to consecutive overlapping sequence of YefM protein were arrayed on a cellulose membrane matrix and covalently bound to a Whatman 50 cellulose support (Whatman). Approximately 50 µg of soluble GST-YoeB proteins were examined for their selective peptide binding ability, on the basis of YefM-YoeB putative interaction. In the case of a low stringency binding procedure, membrane was briefly washed in 100% ethanol, three times washed with Tris-buffered saline (TBS; 50 mM Tris-HCl pH 7.5, 150 mM NaCl), and then blocked for 4 hours using 5% (w/v) non-fat milk in TBS. Next, membrane was washed three times in TBS + 0.1% (v/v) tween 20 (TBS-T), and incubated for 14 hours with 10 ml GST-YoeB solution at slow shacking at 4 °C. Subsequently, the membrane was washed once in TBS-T. Membrane was then added with 10 ml TBS, mouse anti-GST antibody and horseradish peroxidase conjugated goat anti-mouse antibody in the appropriate titers. After 1 hour incubation at room temperature, membrane was briefly washed with TBS-T and TBS. When high stringency binding procedure was performed, washing steps were extensive and multiple. Moreover, the blocking solution washing step was reduced to a single brief wash. Bound GST-YoeB proteins were detected through the enhanced chemiluminescence reaction after an exposure to a sensitive film.

Results

Identification of the yefM-yoeB system genes

The YefM protein of *E. coli* was suggested to be homologous to the Phd protein [23], and similar to the Phd antitoxin was considered to serve as the antitoxin partner of a YoeB toxin. However, this homology is very low and, in fact not statistically significant ($E = 18$, according to pairwise BLAST analysis). This is still very intriguing since the Phd protein appears to have unique structural properties and shows no clear homology to any other proteins. In order to justify the suggested 'YefM-Phd protein family' term [23], systematic exploration of YefM and Phd protein sequences is essentially required. Homologues of YefM were demonstrated to reside on the *Francisella tularensis* plasmid pFNL10 [23] and on a multidrug resistance plasmid identified in a clinical isolate of *Enterococcus faecium* [24]. The existence of homologues of YefM and YoeB protein in bacterial chromosomes was also suggested [24]. However, many unpaired YefM and YoeB homologues were presented [24], indicating a methodical YefM and YoeB homologues pairing is required in order to verify their authenticity as a functional module. Therefore, we used a pair-constrained homology search. In this search, a combination of the values of homology (albeit low) for both putative toxin and antitoxin taken together with their chromosomal organization was taken into account. Only pairs of proteins that revealed paradigmatic TA genetic organization, in which the physical distance between the pair of proteins is less than 100 bp, were regarded as putative TA systems. The resulting findings are showed in Fig. 1.

In the view of this homology analysis, it became clear that a subset of the YefM homologue sequences that are highly similar to Phd are located adjacently to

prophage P1 Doc protein homologues, instead of YoeB (Fig. 1 and Fig. 2A). Therefore, we relate these sequences as hypothetical *phd* genes. This group includes translations of genomic sequences from *S. typhimurium*, *K. pneumonia*, and *Y. enterocolitica*. Those bacteria are actually closer in sequence to Phd (with an *E* value of 2×10^{-9}, 7×10^{-9} and 2×10^{-4}, respectively) than to *E. coli* YefM ($E = 2 \times 10^{-4}$, 3×10^{-4}, 0.8, respectively). Anyhow, these two systems may exist together: the *Y. enterocolitica* bacterium includes both YefM-YoeB and Phd-Doc homologue sequences on its genome (see Fig. 1 and 2).

Alignment of all of the homologous translated sequences was conducted in order to estimate their rate of conservation. YefM homologues alignment (Fig. 2A) consists 29 different homologues, in addition to the Phd protein sequence of phage P1 (last sequence). The toxins alignment (Fig. 2B) is divided into two sections: upper panel includes the YoeB homologues, consisting of 26 different sequences, and the lower panel includes the Doc homologues alignment, consisting of 3 different Doc homologues in addition to the Doc protein sequence of the phage P1 itself. YoeB and Doc homologues cannot be engaged into a reliable alignment due to their far diverse sequences.

The yefM-yoeB genes act as a toxin-antitoxin system

In order to examine the toxic and antitoxic effect that the expressed proteins have on the cell, YefM and YoeB were overexpressed separately and together as an operon using the pBAD-TOPO plasmid. *E. coli* TOP10 strains, carrying the plasmids, were grown in LB medium and 0.2% arabinose added at time zero. Significant effect was observed in these bacteria (Fig. 3A-C). The over-expression of the putative toxin,

YoeB, inhibited the bacterial growth to maximum OD_{600} of approximately 0.15 (Fig. 3B). Over expression of both YefM and YoeB as an operon abolished this toxic effect, indicating toxin-antitoxin relationship between YoeB and YefM (Fig. 3C), as accepted [24]. Surprisingly, over-expression of YefM alone had displayed similar effect on cell growth as YoeB did (Fig. 3A). Same results had been witnessed when cells expressing the system genes were induced during the logarithmic growth stages (Fig. 3D-F): 0.2% arabinose was added to the different cultures at the time they reached OD_{600} of approximately 0.45. In the cases of YefM or YoeB expression, absolute growth inhibition had been observed after less than 1 hour (Fig. 3D,E) as cells reached approximately 0.7 OD_{600}, while the expression of both genes together enabled normal growth (Fig. 3F).

To confirm that the YefM is an actual antitoxin, we tested the colony formation capability of each of the clones at decreasing expression levels (Fig. 3G). On the whole, YefM-expressing clones have consistently demonstrated certain degree of growth in all arabinose concentrations, whereas *yoeB* clones did not form colonies at most concentrations. Moreover, in the presence of 0.005% arabinose, growth of the *yoeB* clone was disabled while the *yefM* clone still demonstrated clear growth, indicating that YoeB is a real toxin while YefM displays toxicity upon high expression levels.

Biophysical characterization of YefM– YefM is natively unfolded

YefM was purified as described in 'Experimental Procedures' section, either by performing gel filtration (obtaining approximately 0.1 mg/ml), or by boiling GST and YefM proteins subsequent to factor Xa cleavage (approximately 0.35 mg/ml).

The far UV circular dichroism (CD) spectra of the purified YefM protein (in both purification methods) at increasing temperatures (25, 37, 42 °C) show a typical random-coil pattern with a minimum in the vicinity of 200 nm [25], with only slight changes in spectra due to temperature increase (Fig. 4A). FTIR spectroscopy also indicates for YefM protein being random-coil structured (Fig. 4B). The FTIR spectrum of the purified YefM (room temperature) showed a transmittance minimum at 1643 cm^{-1} relating to random-coil structure [26].

Thermal denaturation experiment (Fig. 4C) approves that YefM keeps a consistent predominant random-coil structure at all temperature range, as continuous temperature increase of the YefM sample from 2 °C to 80 °C did not significantly shift the CD ellipticity at 222 nm or at 217 nm (wavelengths specifying for maximum CD ellipticity of α-helix and β-sheet structures, respectively), implying structure remained unchanged. Another support for the natively unfolded state of YefM comes from its extraordinary solubility during boiling (Fig. 4D).

Amino acid composition of YefM family proteins

In order to visualize differences between amino acid composition of the YefM proteins and the general amino acid composition, and to gain further insight into the role of sequence in providing disorder characteristics, we have compared the general occurrence of each amino acid in relation to its mean occurrence in YefM proteins. As shown in Fig. 5A, YefM family proteins are considerably enriched in M and E (30-50%), and substantially depleted in W, C, P, F and G (> 50%). The obtained results for these amino acids are much significant, with p-value<0.001, as determined by a

one-sample *t* test. All other amino acids do not display significant enrichment or depletion from the general occurrence of amino acids.

Hydrophobicity-charge relationships in the YefM family proteins

A comparative study that was published by Uversky et al. [3] well demonstrates that it is possible to predict whether a given sequence encodes a folded or natively unfolded protein by a two dimensional plot of the overall hydrophobicity and the net charge of the studied proteins. In order to assess whether the hydrophobicity-charge properties of the YefM family proteins correlate with those previous findings, we have examined these relationships for YefM, Phd and their homologue sequences as described previously [3] (Fig. 5B). Unexpectedly, the YefM-Phd family proteins were found to be mostly localized within the defined 'folded region' of the plot. Interestingly, the localization of Phd protein and its homologues is indistinguishable from the YefM homologues.

Identification of YefM Recognition Determinant

On the basis of YefM's natively unfolded structure, we assumed a linear determinant rather than a conformational one to be recognized by its toxin partner. To identify this determinant in the YefM sequence, we have designed an array consisting of 41 overlapping tridecamer peptides corresponding to amino acids residues 1-12 up to 80-92 of the whole YefM sequence in successive order with 2 amino-acids shifts (Fig. 6A), synthesized on a cellulose membrane matrix. The YefM fragments capable of binding GST-YoeB fusion were identified by immunobloting. Using a low stringency procedure to obtain maximum putative interaction sites, we have identified three such regions. As seen in Fig. 6A, first region included three tridecamer peptites ($YefM_{11}$-

$_{23}$–YefM$_{15-27}$) in decreasing binding capacity, including the sequence RTISYSEARQNLSATMM (underlined sequence represents major bound site); second region included the single YefM$_{33-45}$ peptide sequence- APILITRQNGEAC; the third region comprised the two YefM$_{75-87}$ and YefM$_{77-89}$ peptides, which cover the MDSIDSLKSGKGTEKD Sequence.

In order to verify our results, we used a second peptides array membrane comprising those regions with the intention of performing a high-resolution analysis of the putative binding sites (Fig. 6B). We used a high stringency procedure (see 'Experimental Procedures') to minimize unspecific binding of the GST-YoeB fusion protein or antibodies. The examined sites were extended to include YefM$_{8-31}$ as the first region; YefM$_{29-48}$ as the second region; and YefM$_{72-92}$ as the third region. The shift between each arrayed tridecamer peptide was reduced to a single amino acid. Out of all examined regions, the YefM$_{11-23}$ peptide (RTISYSEARQNLS) was detected as the best YoeB binding sequence.

The arginine in position 19 is essential for YefM-YoeB interaction

Alongside with the verification of the major binding sequence, we tried to detect a single amino acid that would be crucial for YefM-YoeB interaction. The identified binding sequence is rather conserved through the YefM-Phd proteins family. However, two amino acids are notably conserved within: arginine (position 19) and leucine (position 22), as seen in Fig. 2A. We have examined the binding capability of a GST-YoeB fusion to a cellulose membrane array using tridecamer peptides corresponding to the YefM$_{11-23}$ sequence, containing R19 or L22 replacements to alanine or glycine (Fig. 6C). While L22 to A and L22 to G replacements only

attenuated the binding of YoeB, replacement of R19 to A or G totally interrupted the binding, suggesting that the arginine in position 19 is essential for the binding of the YoeB toxin.

Discussion

Non-native protein structures attract an increasing degree of intention due to their abundance on the one hand, and the lack of understanding of their physiological significance on the other hand. Identification of distinct families of natively unfolded proteins, understanding of their conservation on the structural level, and understanding of their physiological role is therefore of high importance. Here, using a combination of bioinformatics, biophysical and physiological analysis, we define a new family of natively unfolded proteins, the YefM-Phd family. Using a pair-constrained bioinformatic approach, we were clearly able to demonstrate that members of the family are present in a large number of bacteria. While the level of homology within the antitoxins family is relatively low (Fig. 2A), we were surprised to find Phd homologues that share higher percentage of homology to YefM than Phd does (*Y. enterocolitica*, *K. pneumoniae* and *S. typhimurium*). Although YefM and Phd proteins share very low sequence homology, the key feature that the proteins share is the natively unfolded state at physiological temperatures (Fig. 4, [11, 27]). Since both Phd-Doc [12] and YefM-YoeB (Fig. 3) are proved be functional TA systems, these findings may suggest that Phd and YefM antitoxins have evolved from a common ancestor system and that at a certain point in the past the antitoxin may have branched out to establish new TA systems consisted of different toxins.

Interestingly, the level of homology within the YoeB family (Fig. 2B) appears to be significantly higher as compared to the YefM family of proteins (Fig. 2A). The level of conservation observed with the YoeB proteins is well consistent with a toxic activity that explicitly targets a specific cellular determinants and that requires a well-defined fold such as a key-lock or induced fit recognition. On the other hand, the low degree of conservation of the extended YefM-Phd family is consistent with a protein missing a clear structural recognition and/or catalytic activity that otherwise requires a defined configuration. It is important to note that YefM and Phd proteins could be irregularly conjugated to a Doc-like or YoeB-like toxins, two families of toxins that could not be aligned and do not share any substantial homology. It is more consistent with a family of protein that is essentially designed to be recognized as a damaged protein and does not represent an interactive or catalytic scaffold. Moreover, the relatively small area of YefM that shows the highest level of conservation was identified to include the target of linear recognition by the YoeB protein (Fig. 6).

Physiological assays have verified that the *yoeB* gene encodes a toxin that is lethal or inhibitory to host cells, and that yefM encodes an antitoxin that prevents the lethal action of the toxin (Fig. 3, [24]). Unexpectedly, upon overexpression YefM inhibited the bacterial growth. However, the dose-dependant behavior of toxicity may suggest that it is an artefact of overexpression rather than a true physiological phenomenon (Fig. 3G).

It is hypothesized that the proteolytical stability difference of the TA system components arises from their thermodynamic stability difference. YefM strongly supports this hypothesis as it was demonstrated to be a natively unfolded protein.

Furthermore, among all structurally described antitoxins – Phd of P1 [11,27], ParD of RK2/RP4 [28], CcdA of F [29] and ε of pSM19035 [31], YefM is the most unstable protein. One of the general structural characteristics of a natively unfolded protein is the lack in secondary structures. At 37 °C, the Phd antitoxin seems to be in a largely unfolded, random-coil conformation as well [11]. However, at 4 °C or at 37 °C in the presence of trimethylamine *N*-oxide (TMAO) chemical chaperon, Phd folds into an ordered protein containing approximately 45% α-helix. Analysis of YefM's far-UV CD spectra yields low content of ordered secondary structure (α-helices and β-sheets) and does not change even at low temperature of 2 °C (Fig. 4A,C) or the addition of "chemical chaperons" (data not shown). YefM was also confirmed to be random-coil by FTIR analysis (Fig. 4B). Additional substantiation for YefM being a most unstructured protein comes from its unusual resistance to aggregation upon boiling (Fig. 4D), which is consistent with a lack of secondary structure elements that mediate aggregate formation through intermolecular association (see Fig. 4D).

It was recently suggested that the relations between sequence and disorder proteins include amino acid compositional bias and high-predicted flexibility [6,31]. According to this study, it was demonstrated that natively unfolded proteins are substantially depleted in Trp, Cys, Phe, Ile, Tyr, Val, Leu and Asn (amino-acid presented in three letters code), and substantially enriched in Ala, Arg, Gly, Gln, Ser, Pro, Glu and Lys. Indeed, we found that the same amino acid compositional bias is valid when comparing the occurrence of the above disordered sequences (using 'ALL-disorder' sequences database [31]) with the general occurrence of amino acids [22] (data not shown). In addition, the depleted amino acids were shown to correspond to low flexibility residues, while the enriched amino acids corresponded to

high flexibility ones [6]. The flexibility ranking is based on a scale developed by Vihinen et al. [32], and reflects the propensity of a given residue to be buried or exposed (i.e. low or high flexibility, respectively) in the crystal structure of globular proteins. However, the amino acid composition of the natively unfolded YefM family proteins is rather different (Fig. 5A). While both the studied disordered proteins and the YefM family proteins are significant depleted in Trp, Cys and Phe, the YefM proteins are further depleted in Gly and Pro – amino acids considered as disorder promoting [6,33]. Moreover, Glu is the sole amino acid seems to be significantly enriched in both. Noteworthy, the most rigid residues (Trp, Cys and Phe) remained depleted in both surveys, insinuating essential importance in the absence of core forming side-chains in the coding of intrinsically disordered sequences.

Recent comparative studies suggested that it is possible to predict whether a given sequence encodes a folded or natively unfolded protein [3-5]. This suggests that a natively unfolded protein must possess the combination of low mean hydrophobicity and relatively high net charge under physiological conditions. However, the majority of the YefM family proteins do not correlate with this determination, including YefM and Phd proteins (Fig. 5B). Obviously, this result is coupled with the unique amino acid compositional bias of the YefM family proteins mentioned above, which does not fit the established characteristics of disordered sequences. The relative lack in high flexibility side-chains (e.g. Lys, Pro, Gly, Ser and Gln) together with an insufficient depletion in hydrophobic rigid side-chains (e.g. Ile, Tyr, Val, and Leu), account for the relatively low net charge and rather high overall hydrophobicity that characterize the YefM family. Furthermore, in the case of the YefM family proteins, we propose that the lack of aromatic residues, rather than hydrophobic, maintains the

disordered state of YefM. As seen in Fig. 5A, the depletion in the aromatic residues Phe and Trp, unlike other hydrophobic residues, is conserved through the YefM family. The lack of aromatic moieties is consistent with the lack of organized and packed hydrophobic core.

As discussed in the introduction section, TA system may serve as an excellent target for antibacterial agent. One approach is to prevent the toxin and antitoxin components from interacting *in vivo*, which would trigger their inhibitory (or lethal) effect on cell growth. As we have identified the molecular recognition sequence within the YefM protein (Fig. 6), we intend to use this information for the design of agents that will affect the YefM-YoeB interaction.

References

1. Anfinsen, C.B. (1973) *Science* **181**, 223-230.
2. Schweers, O., Schönbrunn-Hanebeck, E., Marx, A. and Mandelkow, E. (1994) *J. Biol. Chem.* **269**, 24290-24297.
3. Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) *Proteins Struct. Funct. Genet.* **41**, 415-427.
4. Uversky, V.N. (2002) *Prot. Sci.* **11**, 739-756.
5. Uversky, V.N. (2002) *Eur. J. Biochem.* **269**, 2-12.
6. Dunker, A.K., Lawson, J.D., Brown, C.J. Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C.-H., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. and Obradović, Z. (2001) *J. Mol. Graph. Model.* **19**, 26-59.
7. Dobson, C.M. (1999) *Trends Biochem. Sci.* **24**, 329-332.
8. Rochet, J.C. and Lansbury, P.T.Jr. (2000) *Curr. Opin. Struct. Biol.* **10**, 60-68.
9. Gazit, E. (2002) *Curr. Med. Chem.* **9**, 1725-1735.
10. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradović, Z. (2002) *Biochemistry* **41**, 6573-6582.
11. Gazit, E. and Sauer, R.T. (1999) *J. Biol. Chem.* **274**, 2652-2657.
12. Lehnherr, H., Maguin, E., Jafri, S. and Yarmolinsky, M.B. (1993) *J. Mol. Biol.* **233**, 414-428.
13. Lehnherr, H. and Yarmolinsky, M.B. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3274-3277.
14. Christensen, S.K. and Gerdes, K. (2003) *Mol. Microbiol.* **48**, 1389-1400.

15. Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K. and Ehrenberg, M. (2003) *Cell* **112**, 131-140.

16. Pedersen, K., Christensen, S.K. and Gerdes, K. (2002) *Mol. Microbiol.* **45**, 501-510.

17. Christensen, S.K., Mikkelsen, M., Pedersen, K. and Gerdes, K. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14328-14333.

18. Gerdes, K. (2000) *J. Bacteriol.* **182**, 561-572.

19. Aizenman, E., Engelberg-Kulka, H. and Glaser, G. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6059-6063.

20. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nucleïc Acids Res.* **25**, 3389-3402.

21. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Nucleic Acids Res* **22**, 4673-4680.

22. http://prowl.rockefeller.edu/aainfo/masses.htm.

23. Pomerantsev, A.P., Golovliov, I.R., Ohara, Y., Mokrievich, A.N., Obuchi, M., Norqvist, A., Kuoppa, K. and Pavlov, V.M. (2001) *Plasmid* **46**, 210-222.

24. Grady, R. and Hayes, F. (2003) *Mol. Microbiol.* **47**, 1419-1432.

25. Jenness, D.D., Sprecher, C.A. and Johnson, W.C. (1976) *Biopolymers* **15**, 513-521.

26. Haris, P.I. and Chapman, D. (1995) *Biopolymers* **37**, 251-263.

27. Gazit, E. and Sauer, R.T. (1999) *J. Biol. Chem.* **274**, 16813-16818.

28. Oberer, M., Lindner, H., Glatter, O., Kratky, C. and Keller, W. (1999) *Biol. Chem.* **380**, 1413-1420.

29. Thi, M.H.D., Messens, J., Wyns, L. and Backmann, J. (2000) *J. Mol. Biol.* **299**, 1373-1386.

30. Camacho, A.G., Misselwitz, R., Behlka, J., Ayora, S., Welfle, K., Meinhart, A., Lara, B., Saenger, W., Welfle, H. and Alonso, J.C. (2002) *Biol. Chem.* **383**, 1701-1713.

31. Romero, P., Obradović, Z., Li, X., Garner, E.C., Brown, C.J. and Dunker, K. (2001) *Proteins Struct. Funct. Genet.* **42**, 38-48.

32. Vihinen, M., Torkkila, E. and Riikonen, P. (1994) *Proteins* **19**, 141-149.

33. Radivojac, P., Obradović, Z., Brown, C.J. and Dunker, A.K. (2003) *Pac. Symp. Biocomput.* **8**, 216-227.

Figure Legends

Fig. 1: Comparative genetic organization of the *yefM* and *yoeB* proteins family. A graphic representation of the size and the physical distance (in base-pairs) between TA coding sequences. The black half-ovals represent homologue sequences of YefM ('antitoxins'), and the gray half-ovals represent homologue sequences of YoeB ('toxins'). Homologue sequences of the Doc protein are represented as sharp gray arrowheads, indicating that their YefM-like antitoxins are regarded as Phd homologues. Missing gi numbers indicates unannotated ORFs. In the case of *F. tularensis* plasmid pFNL10, the *yefM* and *yoeB* are regarded as orf5 and orf4, respectively.

Fig. 2: YefM and YoeB sequence alignments. A. Multiple sequence alignment of yefM proteins family. Alignment list includes 30 sequences from 25 different bacteria (different homologues in the same bacteria are indicated in alphabetical order). Residues that are similar in ≥ 80% sequences are colored in dark blue background. Residues that are similar in ≥ 60% and ≥ 40% are colored in medium and light blue background, respectively. Identity percentage is based on BLOSUM62 matrix values.

B. Multiple sequence alignment of YoeB proteins family. Upper alignment list includes 26 sequences from 22 different bacteria, all showing homology to YoeB protein. Lower list include 3 Doc homologues protein sequences. YoeB and Doc sequences do not align (*E*-value > 10^6), so as their homologues. The alignment was generated and is colored as described in Fig. 2A.

Fig. 3: Demonstration of antitoxin and toxin activity of YefM and YoeB. *E. coli* strain TOP10 carrying one of the pBAD-TOPO vectors expressing YefM **(A, D)**; YoeB **(B, E)**; or YefM-YoeB together as an operon **(C, F)**, were grown in LB-Amp medium at 37 °C. Transcription of the respective genes was induced by the addition of 0.2% arabinose (full circles) at two different growth phases: stationary- at time zero **(A-C)**, and logarithmic- when cultures reached OD_{600} of 0.45 **(D-F)**. In Parallel, equal culture volumes were added with 0.2% Glucose as a negative control (open circles). **G.** The effect of overexpressing YefM, YoeB, or YefM-YoeB together in a TOP10 strain. Dropouts of the different clones (as indicated) were plated on arabinose gradient plates in 10-fold dilutions and incubated for 20 hours at 37°C. The arabinose gradient plates are in the following order (top to bottom): 0%, 0.0005%, 0.005%, 0.02%, 0.05%, 0.1%, and 0.2%. Plates missing L-arabinose were added with 0.2% glucose.

Fig. 4: YefM protein is natively unfolded. A. Circular dichroism spectra. CD spectra of YefM at 25 °C (- - -), 37 °C (——), and 42 °C (— —) in PBS pH 7.3. Spectra pattern corresponds to random coil structures. Same protein sample was incubated at the different temperatures. **B.** Fourier Transform Infrared spectra of YefM protein. Minimum transmittance at wavenumber of 1643 cm^{-1} indicates a random coil structure of the sample. **C.** Thermal Denaturation between 2 °C and 80 °C. Thermal stability was determined by monitoring CD ellipticity at 217 nm (triangles) and 222 nm (circles) as a function of temperature. **D.** YefM remains soluble through boiling. Left lane: YefM and GST proteins following factor Xa cleavage reaction. Right lane: supernatant content after 10 minutes boiling followed by 10 minutes centrifuging.

Fig. 5: Analysis of the physicochemical properties of the identified proteins. A. YefM amino acid occurrence relative to the general amino acid occurrence [22], given by $(P_{Mi} - P_{Gi}) / P_{Gi}$. Error bars represent the standard deviations. Significance of difference between the antitoxins' amino acids mean occurrences and the general occurrences designated by *, indicates P<0.001 as determined by one sample *t* test. The amino acids are arranged according to residue flexibility [32], with increasing flexibility to the right. **B.** Comparison of the mean net charge and the mean hydrophobicity for the YefM (circles) and the Phd (triangles) proteins family. The solid line represent the border between natively unfolded proteins (upper left) and folded proteins (bottom right) calculated using the equation <R>=2.785<H>-1.151, as was proposed by Uversky and coauthors [3]. The YefM protein (gray circle), Phd protein (gray triangle) and their homologues are mostly localized in the 'folded' region. Mean net charge and mean hydrophobicity were calculated as described in [3].

Fig. 6: Identification of the YoeB binding sequence in the YefM protein using a peptide array. A. 41 Tridecamer peptides corresponding to consecutive overlapping sequences of 92 a.a. YefM protein (two amino-acids shift between peptides) were arrayed on a membrane. GST-YoeB bound to the membrane was detected. **B.** Tridecamer peptides corresponding to consecutive overlapping sequences of $YefM_{8-31}$, $YefM_{29-48}$, and $YefM_{72-92}$ (single amino-acid shift between peptide) were arrayed on a membrane and analyzed for GST-YoeB binding. **C.** Tridecamer peptides corresponding to YefM-YoeB recognition sequence with R19 and L22 replacements were analyzed for GST-YoeB binding. No GST-YoeB binding could be detected to R19A or R19G tridecamer peptide.

Claims:

1. A bioinformatics method for finding a plurality of remotely homologous bacterial protein sequences based on a plurality of known bacterial protein sequences, said plurality of remotely homologous and known bacterial proteins being expressed in vivo in equimolar ratios, the method comprising bioinformatically searching at least one bacterial sequence database for a plurality of sequences remotely homologous to said plurality of known bacterial protein sequences, said plurality of sequences remotely homologous to said plurality of known bacterial protein sequences residing on a genome of a bacterial species in a distance no greater than a predetermined value.

2. The method of claim 1, wherein said bacterial protein sequences are pairs of toxins and anti-toxins.

3. A method of defining or isolating a determinant portion of a toxin capable of binding an anti-toxin, the method comprising interacting peptides derived from said toxin with said anti-toxin and monitoring an interaction between said peptides derived from said toxin with said anti-toxin, thereby defining or isolating said determinant portion of said toxin capable of binding an anti-toxin.

4. A process of drug development based on a determinant defined or isolated by the method of claim 3, comprising obtaining analogs of said determinant and determining for said analogs interaction capability with said anti-toxin.

5. A determinant defined or isolated by the method of claim 3.

6. A drug developed by the process of claim 4.

7. A pharmaceutical composition comprising the drug of claim 6.

8. A method of treating bacterial infection comprising administering to a subject in need thereof the drug of claim 6, the pharmaceutical composition of claim 7 or a determinant isolated by the method of claim 3.

Fig. 1

Organism	gi number	Gene size & distance	Dis.(bp)
A. actinomycetemcomitans			-1
A. tumefaciens	17740473;17740472		-4
B. cepacia			58
C. burnetii			5
E. coli	1788328 ;3915522		-1
E. faecium pRUM plasmid	28849326;28849325		-5
F. tularensis pFNL10 plasmid	4325240		-1
K. pneumoniae			57
M. bovis			65
M. tuberculosis	15610493;7477590		-4
N. europaea A	;22954686		-1
N. europaea B	22955251;		10
N. europaea C			47
Nostoc sp. PCC 7120	17135150;17135149		-11
P. fluorescens	:23062981		2
P. putida	24984546;24984545		21
P. syringae	28850803;28850802		29
R. conorii	15892044;		80
S. typhimurium	16422119;16422118		-1
S. aureus	14248182;15925397		-1
S. pneumoniae	14973237;14973236		3
S. coelicolor	4469319 ;4469318		-1
S. viridochromogenes	15077435;15077436		-4
Synechocystis sp. PCC 7942	22002543;22002544		-5
Synechocystis sp. PCC 6803 A	16331058;		-4
Synechocystis sp. PCC 6803 B	16330633;		-1
T. ferrooxidans			80
Y. enterocolitica A			5
Y. enterocolitica B			56

Fig. 2A

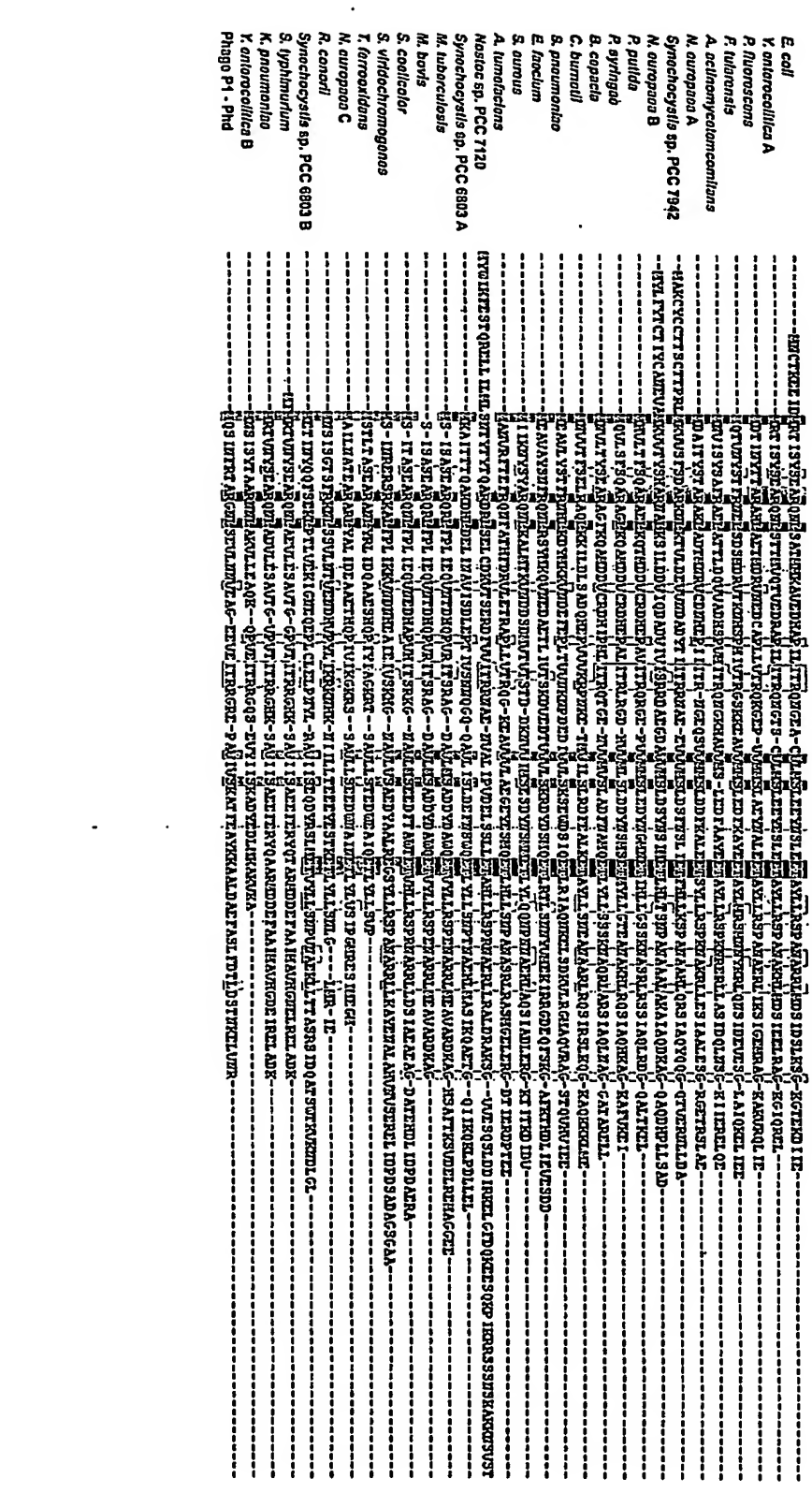

E. coli
Y. enterocolitica A
P. fluorescens
P. fulvomidis
A. actinomycetemcomitans
N. europaea A
Synechocystis sp. PCC 7942
N. europaea B
P. putida
P. syringae
B. capacia
C. burnetii
S. pneumoniae
E. faecium
S. aureus
A. tumefaciens
Nostoc sp. PCC 7120
Synechocystis sp. PCC 6803 A
M. tuberculosis
M. bovis
S. coelicolor
S. viridochromogenes
T. ferrooxidans
N. europaea C
R. conorii
Synechocystis sp. PCC 6803 B
S. typhimurium
K. pneumoniae
Y. enterocolitica B
Phage P1 - Phd

Fig. 2B

P. putida
P. syringae
A. tumefaciens
B. cepacia
E. coli
Y. enterocolitica A
N. europaea A
P. fluorescens
Synechocystis sp. PCC 6803 A
S. aureus
F. tularensis
A. actinomycetemcomitans
S. pneumoniae
E. faecium
R. conori
N. europaea B
Synechocystis sp. PCC 7942
C. burnetii
M. tuberculosis
M. bovis
S. coelicolor
S. viridochromogenes
Nostoc sp. PCC 7120
Synechocystis sp. PCC 6803 B
T. ferrooxidans
N. europaea C
K. pneumoniae
S. typhimurium
Y. enterocolitica B
Phage P1 - Doc

Fig. 3

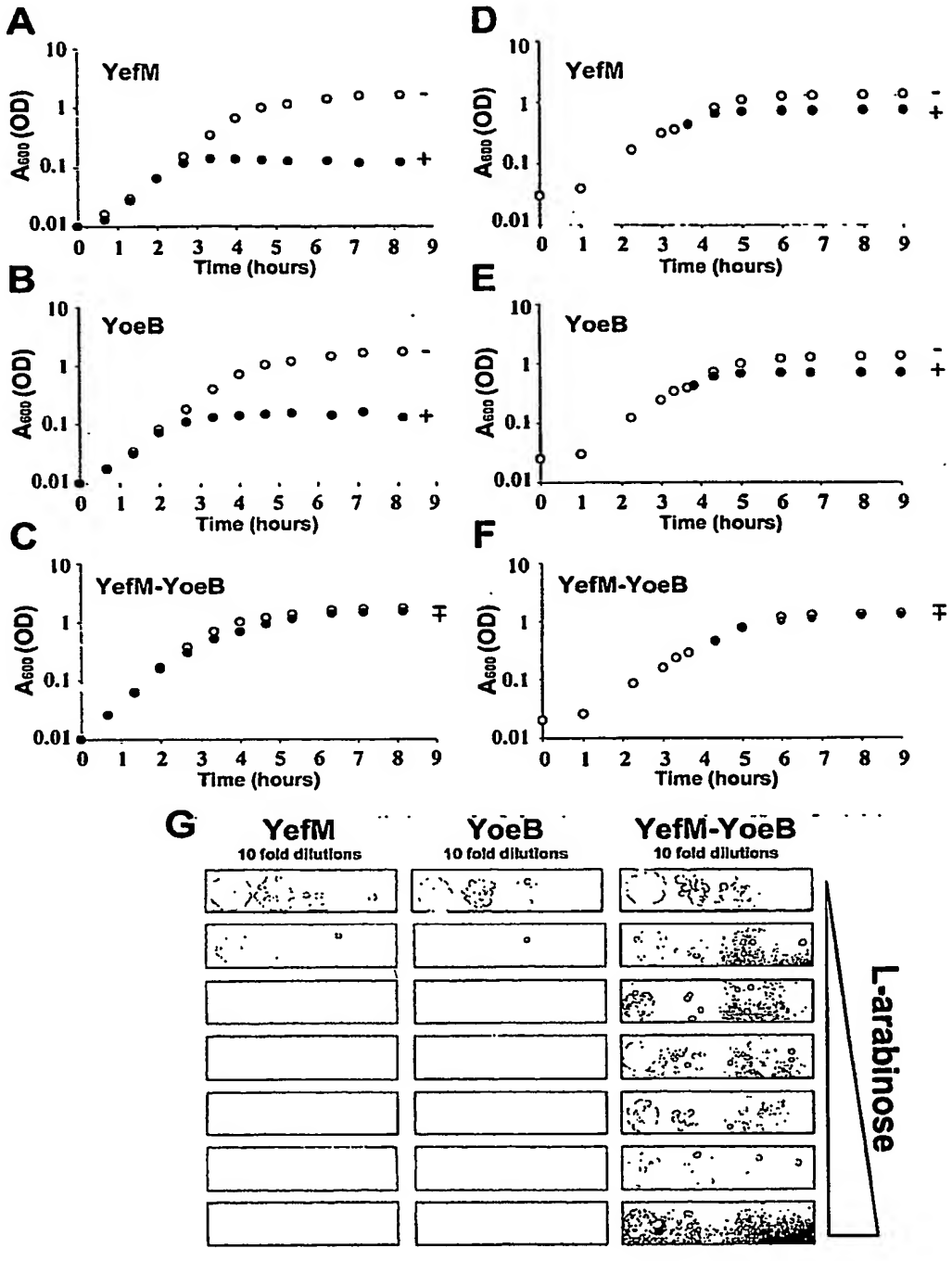

A
YefM
B
YoeB
C
YefM-YoeB
D
YefM
E
YoeB
F
YefM-YoeB
A600 (OD)
Time (hours)
G
YefM
YoeB
YefM-YoeB
10 fold dilutions
L-arabinose

Fig. 4

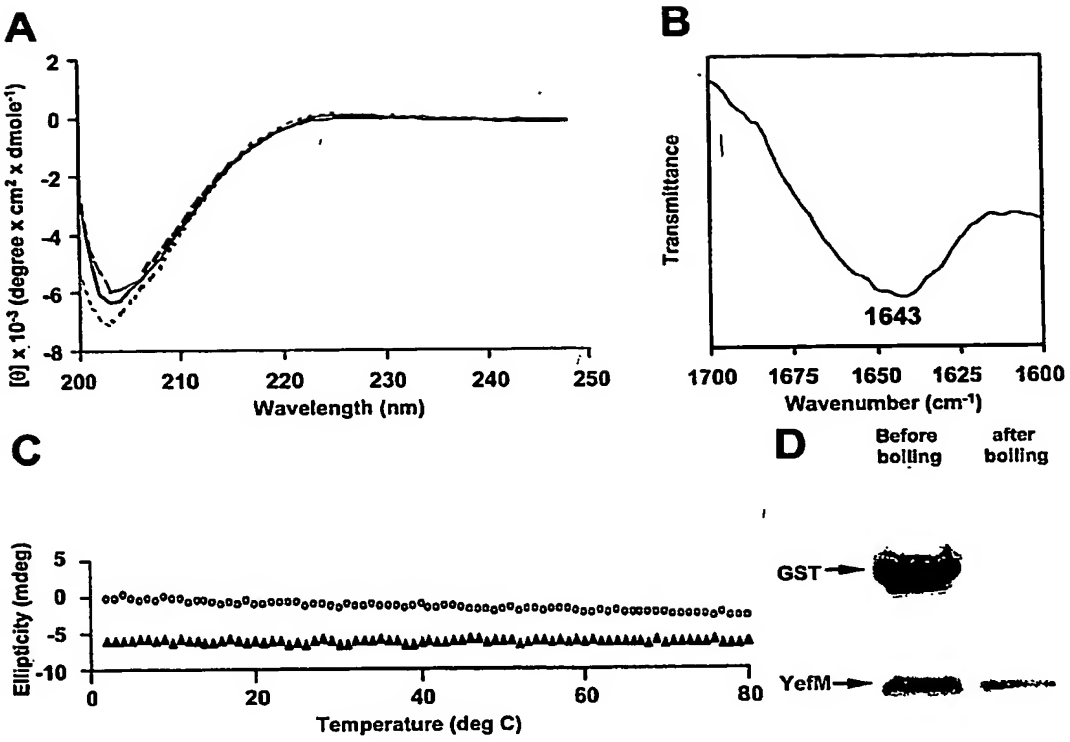

A
[θ] x 10⁻³ (degree x cm² x dmole⁻¹)
2
0
-2
-4
-6
-8
200
210
220
230
240
250
Wavelength (nm)
B
Transmittance
1643
1700
1675
1650
1625
1600
Wavenumber (cm⁻¹)
C
Ellipticity (mdeg)
5
0
-5
-10
0
20
40
60
80
Temperature (deg C)
D
Before boiling
after boiling
GST
YefM

Fig. 5

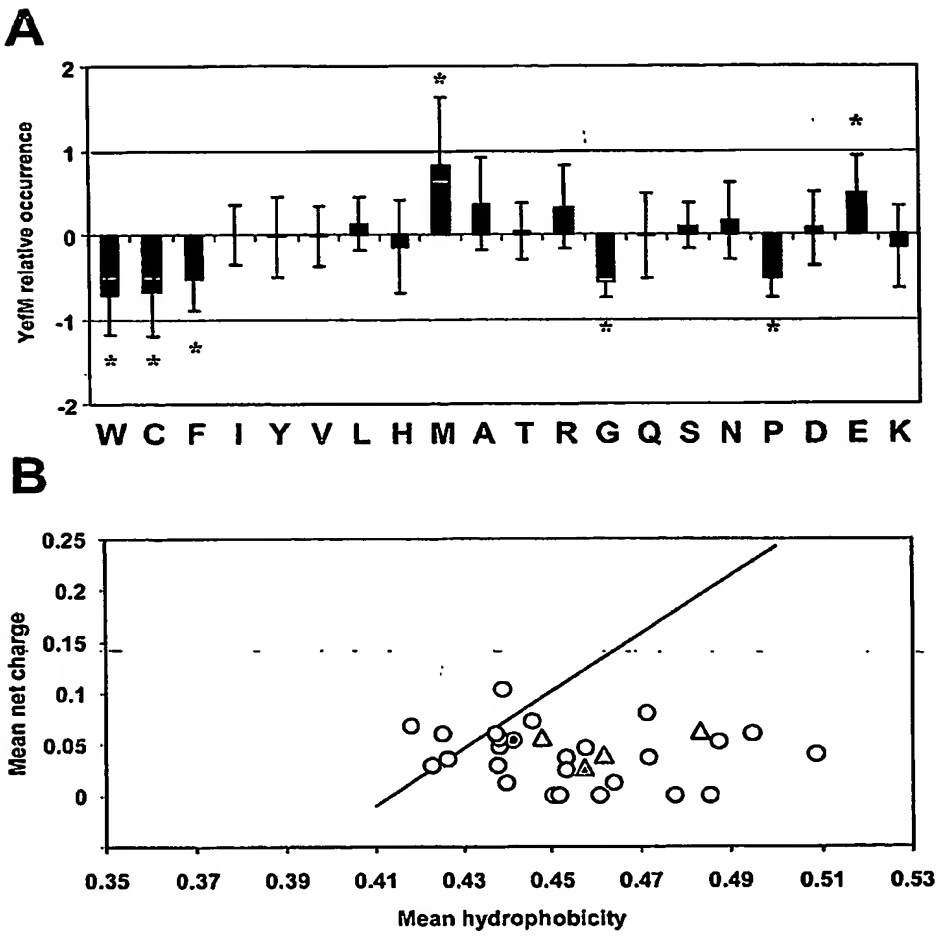
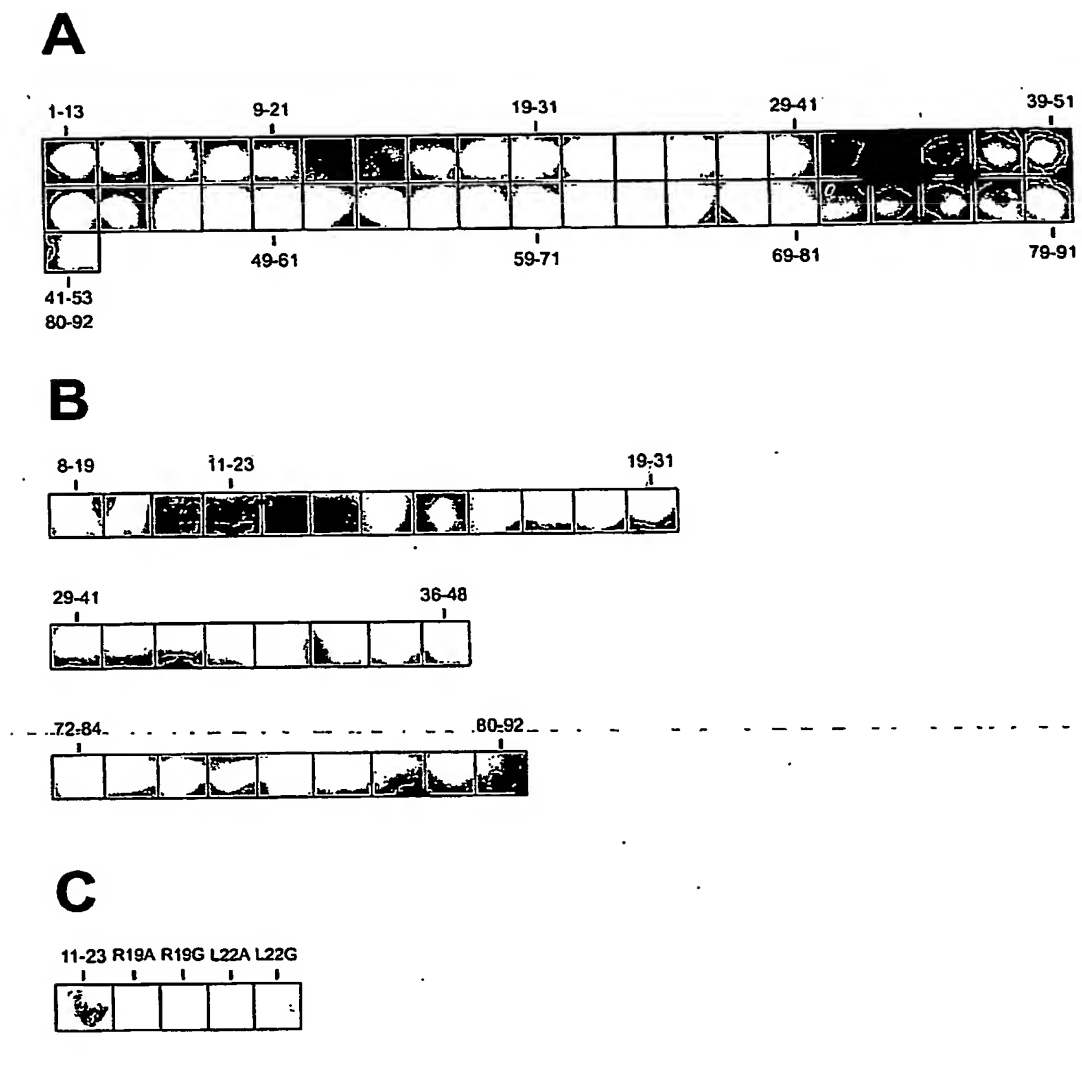

A
YefM relative occurrence
2
1
0
-1
-2
W C F I Y V L H M A T R G Q S N P D E K
B
Mean net charge
0.25
0.2
0.15
0.1
0.05
0
0.35 0.37 0.39 0.41 0.43 0.45 0.47 0.49 0.51 0.53
Mean hydrophobicity

Fig. 6


A
1-13
9-21
19-31
29-41
39-51
49-61
59-71
69-81
79-91
41-53
80-92
B
8-19
11-23
19-31
29-41
36-48
72-84
80-92
C
11-23 R19A R19G L22A L22G